DOCUMENT RESUME

ED 189 174 TM 800 356

AUTHOR Smith, Richard M.: Mitchell, Virginia P.
TITLE An Analysis of Rasch Person Fit Statistics in College Placement Tests.
PUB DATE Apr 80
NOTE 28p.: Paper presented at the Annual Meeting of the National Council on Measurement in Education. (Boston, MA, April 8-10, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests: Computer Oriented Programs: *Difficulty Level: *Goodness of Fit: Guessing (Tests): Item Analysis: *Latent Trait Theory: *Student Placement: *Test Items: *Two Year Colleges
IDENTIFIERS New Jersey College Basic Skills Placement Test: *Rasch Model

ABSTRACT

To analyze test item and person interactions according to the Rasch Model, 822 incoming community college freshman were administered the New Jersey Basic Skills Placement Test. The Logical Relationships and Elementary Algebra subtests were chosen to examine the effect of differing average item difficulty on descriptions of fit statistics and on frequency of person misfit. Three user-defined item groups: sequence, subscores, and difficulty, were chosen to highlight three misfit statistics: working too slowly to finish test, misunderstanding content, and guessing. A fourth misfit statistic, total t test, was also calculated. On the four misfit statistics, the percentage of misfitting students was 25% (Logical Relationships) and 33% (Elementary Algebra). After editing techniques were applied to the item response strings, the algebra subtest became valid for 90% of the misfitting students. The appendix contains samples of the PANAL computer program output for typical cases of misfit, as well as editing suggestions for producing meaningful ability estimates. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED189174

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AN ANALYSIS OF RASCH PERSON FIT STATISTICS IN COLLEGE PLACEMENT TESTS

RICHARD M. SMITH
MERCER COUNTY COMMUNITY COLLEGE

VIRGINIA P. MITCHELL
SYRACUSE UNIVERSITY

March 1980

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R. Smith

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 800356

Paper presented at the annual meeting of the National Council on Measurement in Education

Boston, April 1980

ABSTRACT

The Rasch Model, which produces independent estimates of item difficulty and person ability, will allow for the calculation of the probability of a correct response for any individual on any item. The analysis of the residuals formed by comparing expected outcomes with observed outcomes will provide estimates of the fit of individuals to the model. The test of fit may include all items on the test or various subgroups of interest. These subgroups might include item difficulty, content area, item type, and order of item presentation. This paper analyzes these tests of fit in college placement tests, and recommends appropriate action for misfitting individuals.

AN ANALYSIS OF RASCH PERSON FIT STATISTICS IN COLLEGE PLACEMENT TESTS

The psychometric model proposed by Georg Rasch is the result of his attempt to define the conditions necessary to obtain objective measurement. The Rasch Model differs from other item response theory models in its focus. This model, rather than seeking to include parameters which tend to mold the model to observations, seeks to eliminate from the observations those characteristics which are undersirable in objective measurement. As a result although the Rasch Model is mathematically equivalent to the one-parameter logistic model, there is an underlying philosophic difference in the application of the two models.

Rasch's notion of objective measurement which sought to exclude from the measurement the confounding properties of items, namely guessing, and item characteristic curves with different slopes, leads quite naturally to the investigation of person response patterns. The investigation of these response patterns is in fact the investigation of the fit of the person to the psychometric model. There are numerous types of these item-person interactions which do not lend themselves to improving measurement. These include, but are not limited to, guessing, practice effect, "plodding," item-type interactions, and "sloppiness," or "boredom." It is the detection and analysis of these interactions which will be the primary purpose of this paper.

The process of measurement as defined by Wright and Stone (1979) is as follows:

> Four steps must be taken to use a test to measure a person. First, we must work out a clear idea of the variable we intend to make measures on. Second, we must construct items which are believable realizations of this idea and which elicit signs of it in the behavior of the persons we want to measure. Third, we must demonstrate that these items when taken by suitable persons can lead to results that are consistent with our intentions. Finally, before we can use any person's score as a basis for their measure, we must determine whether or not their particular pattern of responses is, in fact, consistent with our expectations.

Much of the work in traditional test theory and many of the applications of the various item-response models have been concerned with the first three requirements, the construction and analysis of items and groups of items. Little, if any, attention has been given to the final requirement, the analysis of the individual person's pattern of responses. The Rasch Model, which limits measurement to two basic concepts, the ability of the person taking an item and the difficulty of the item, provides an excellent framework for conducting an analysis of an individual's pattern of responses or "person fit." Using this one parameter model, we are able to examine the individual's response pattern for systematic discrepancies by removing the effects of the model from the data. It is the ability to determine the probability of a correct response for any individual on any item given the individual's ability and an item pool calibrated with the Rasch Model that allows one to generate residuals (departures from expected values) by matching these probabilities with actual results.

Mead (1979) showed the effect of various item-person interactions on item characteristic curves. The effects of guessing, speed (practice) and bias were shown through a study of the residuals derived from the difference between the observed and predicted outcomes. Mead and Wright (1977) used this residual to study the effect of inappropriate test-taking behaviors on item characteristic curves. They conclude that fluctuations in item characteristic curves are often due not to a problem with the item but due rather to inappropriate test-taking behaviors. They recommend procedures for removing the effect of this behavior from both the item characteristic curve and the individual ability estimates.

Wright (1977) continued the discussion of person fit and illustrated the properties of guessing and "sleeping" by plotting the residuals against the estimated differences between person abilities and item difficulties (b-d). In the classical guessing case, as the (b-d) difference becomes small, the probability of a correct response becomes small; a correct response due to guessing would, therefore, have a large residual. This residual would become larger as the items become more difficult giving the plot of the residuals for correct responses an upward trend.

Wright and Stone (1979) conclude:

> Therefore, if we are serious in our attempt to measure, we must examine every application to see how well each set of responses corresponds to our model expectations. We must evaluate not only the plausibility of the sample of persons' responses, but also the plausibility of each person's responses to the set of items in his test. To do this we must examine the response of each person to each item to determine whether it is consistent with the general pattern of responses observed.

It was this call for person analysis which led Mead (1978) to write a computer program 'PANAL' which will calculate and display various aspects of a person's fit statistics for calibrated tests. Smith and Mitchell (1979) utilized this program in modifying individual sets of responses on college placement tests.

The purpose of this study was to analyze person fit statistics on various college placement tests, examine the effect of item difficulties on these statistics, improve the techniques for identifying misfitting individuals and discuss techniques for improving the measure of the abilities for these misfitting individuals.

The placement test used for this study was the New Jersey College Basic Skills Placement Test (NJCBSPT). This test was developed in response to a New Jersey Board of Higher Education mandate that the basic skills of all freshmen entering public two-year and four-year colleges in the State of New Jersey be assessed and that the results of this assessment be used for college placement (but not admissions). The NJCBSPT is published by Educational Testing Services and consists of five objective tests and one twenty minute essay. There are three thirty minute objective English tests--Reading Comprehension, Sentence Structure and Logical Relations. There are two thirty-five minute objective math tests--Computation and Elementary Algebra. The English tests were developed from the item pool used by Educational Testing Service to build the California State University and Colleges English Placement Test. The math tests were developed from the item pool used to construct the Descriptive Test of Mathematics Skills.

The initial form of this test (Form 3AJP) was administered to 822 incoming freshmen, at Mercer County Community College, during the Summer of 1978. The results of the test analysis (see Table I) indicated that

one of the five versions (Elementary Algebra) was much more difficult (mean -37% correct) than the other four tests (average mean - 71% correct). This difference was also evident in the State-wide population (N-42,950) where the mean on the Elementary Algebra Test was 47% correct.

The item response strings for 822 incoming freshmen were processed by BICAL (Wright and Mead, 1978), the Rasch Model Calibration Program, to produce item difficulties for the 185 items, and five separate ability estimates (one for each of the five objective tests) for the 822 freshmen. The item difficulties, person ability estimates, and the item response strings were then used as input to the PANAL program to calculate the individual fit statistics.

The results reported in this paper are the results of analyzing two of the tests in the NJCBSPT battery. The Logical Relationships Test and the Elementary Algebra Test were chosen to test the effect of differing average item difficulties on the distribution of fit statistics and the frequency of person misfit. A third sample of randomly generated responses were analyzed to test the frequency of detecting existing misfit through this analysis as well as to determine a general residual pattern for guessing against which the two real samples were compared.

Wright and Stone (1979) provide a review of the process for developing fit statistics for the Rasch Model. A brief overview of this development follows:

The probability P_{vi} for a given response is defined as:

$$P_{vi} = \exp (b_v - d_i)/[1 + \exp (b_v - d_i)]$$

where b_v is the estimated ability of person v and d_i is the estimated difficulty of item 1.

This probability can be viewed as an estimate of the expected value of any given response. By comparing the expected value of a response with the actual response we can develop a standarized residual.

$$Z_{vi} = (X_{vi} - P_{vi})/[P_{vi}\ (1-P_{vi})]^{\frac{1}{2}}$$

which should be distributed normally with a mean of 0 and a variance of 1.

The square of this residual should be distributed as a chi square with 1 degree of freedom, and is easily calculated in the dichotomous response situation where:

when	$X = 0$	$Z_0^2 = \exp(b-d)$
and when	$X = 1$	$Z_1^2 = \exp(d-b)$

These squared residuals can then be accumulated over all of the person-item interactions. These accumulated Z^2's should be distributed as a chi square with n-1 degree of freedom where n is the number of items.

This accumulated Z^2 can be converted to a t distribution using the following transformation:

$$V = \Sigma Z^2/f$$

$$t = [\ln(V) + V - 1]\ [\ f/8]^{\frac{1}{2}}$$

This statistic has approximately a unit normal distribution.

Four t-statistics were calculated for each individual's responses on these two tests. Summing the squared residuals over all the items on a particular test produced the first of the t-statistics used in this study, the "total t." Three between group t-statistics were also calculated. These user-defined groups of items were selected for incorporation into the PANAL program. Since the Rasch Model is "item freed" it is possible to estimate an individual's ability on any subset of the total items. On the Elementary Algebra test for example, the items were divided into four groups on the basis of the estimated item difficulty. On the basis of these four groups (easy, low middle, high middle, most difficult) four independent estimates of a person's ability were obtained. These independent estimates were then used to calculate

the squared residuals. From these squared residuals within group and between group fit statistics were generated for each group. These statistics can be evaluated as t-distributions with mean of 0 and a standard deviation of 1.

Three user-defined groupings were chosen for this study. They are: (see table 2)

Sequence - Dividing the test into 3(4) sequential groups based on the order of item presentation.

Subscores- Dividing the items according to item content.

Difficulty- Dividing the items into groups based on item difficulty.

These groups were chosen to highlight particular types of individual misfit. High values of the between-group fit statistic on items grouped by difficulty usually illustrates guessing. High values of the between-group fit statistic on items grouped by sequence usually indicates "plodding" (working too slowly to finish the test) or "start-up" (having difficulty on early items). Additionally, in cases where items are grouped by item type or roughly arranged in item difficulty order this fit statistic can often detect item type-person interactions or guessing. High values of the between-group fit statistic on items grouped by content usually indicate a person-content interaction.

Table 3 shows the mean values of the t-statistic calculated for the three analyses. The difference in the difficulty of the Elementary Algebra Test and the Logical Relationships Test appears to make the Logical Relationships Test fit the model slightly better than does the Elementary Algebra Test on all four t-statistics. The analysis of the randomly generated responses indicates a very high mean value for the total test and difficulty sub-groups. Since both of these statistics are extremely sensitive to guessing, the result is not unexpected and should lend support to the assumption that these test-of-fit statistics are sensitive to test-taking behavior not included in the model.

Further analysis of the random responses indicated that the total-t fit statistic identified 64% of the guessing cases (Table 4). In all, seventy-nine percent of the guessing cases had one or more fit statistics

above the critical level, while only 11% of these cases had fit statistics less than 1.0. The choice of the critical fit statistic value of 1.5 is worth noting in light of the assumption that the fit statistic had a mean of zero and a standard deviation of 1. This statistic seems to be sensitive to test length. Since we were dealing with fairly short tests and subgroups, some modification in the critical value of the fit statistic seemed necessary. For short tests it is not uncommon to find individuals with unacceptable response patterns in the 1.2 to 1.8 range. Thus, experience indicates the critical value should be set at 1.5.

Table 5 shows the frequency of misfitting individuals on each of the four misfit statistics. The frequency of misfit on the Elementary Algebra test is high. One third of the students who took this test did not fit the model. It would raise serious consequences in any testing situation if one third of the scores could not be considered valid. The frequency of misfit for the Logical Relationships test and the other three objective tests of like difficulty were similar. The number of misfitting individuals on each test was near 25%. It would be erroneous to attribute the high incidence of misfit on the Elementary Albegra test solely to the difficulty of the items, but there seems to be some reason to suspect that there is a relationship between the difficulty of the test and the occurrence of misfit.

Tables 6 and 7 illustrate the identifiable patterns of misfit on the Elementary Algebra test. Approximately two-thirds of the individuals identified as misfitting fell into the range of one to nine correct answers, while only forty-one percent of all the students taking the test fell into this range. It is, therefore, important to be able to separate those students for whom this score is a valid measure from those for whom the measure is invalid. Of the 366 students whose scores fell into this range 55% (203) were identified as having misfit one or more of the calculated statistics. Since the individual's ability under the Rasch Model is independent of the number or difficulty of the items on individual attempts, various editing techniques were applied to these misfitting students. Of the 259 students who misfit

on the Elementary Algebra test, we were able to identify the misfit and edit the item-response strings for 90% of the misfitting students so as to produce a valid measure. There were only 27 students for whom the results were meaningless. These students would have to be retested to obtain meaningful ability estimates.

Appendix A contains samples of the PANAL output for typical cases of misfit. Included in these samples are suggestions as to how the response strings might be edited to produce more meaningful ability estimates. This editing can only be accomplished within the framework of an item response theory model which allows comparable ability estimates to be generated from dissimilar sets of items.

The implications of these procedures are worth noting. The adoption of the Rasch Model for test scoring and item calibration will provide the test user with a systematic and manageable means of checking the quality of each student's results and of the ongoing utility of items used. This procedure can be incorporated into any automated scoring system. The procedure provides the means of checking all test results for misfit on a variety of user-defined demensions. Very often the procedure can indicate the type of misfit present and the appropriate action can be taken.

TABLE I Test Information

	Reading Comprehension	Sentence Structure	Logical Relationships	Computation	Elementa Algebra
Test Length	30 min	30 min	30 min	35 min	35 min
No of Items	40	35	50	30	30
State Wide (n=42950)					
Test Mean(STD Score)	64.5	64.2	64.7	63.9	65.4
STD Dev(STD Score)	11.6	11.3	11.5	11.3	11.0
Test Mean (Raw Score)	30.6	24.7	38.7	21.2	14.0
Average p+*	.765	.706	.774	.707	.467
College Sample N=822					
Test Mean(STD Score)	63.5	62.5	63.0	61.5	61.1
STD Dev(STD Score)	10.8	10.8	11.1	10.6	9.5
Test Mean (raw Score)	29.9	23.7	37.6	19.81	11.0
Average p+*	.748	.677	.752	.660	.367

*p+ + = The percentage of people responding to an item correctly.

TABLE 2

ITEM CLASSIFICATIONS

	Elementary Algebra	Logical Relations
SEQUENCE		
No. of Groups	3	2
No. of Items In Each Group	10, 10, 10	25, 25
SUB-SCORES		
No. of Groups	2	4
No. of Items In Each Group	18, 12	12, 13, 12, 13
DIFFICULTY		
No. of Groups	4	4
No. of Items In Each Group	6, 8, 11, 5	9, 18, 15, 8

TABLE 3

MEAN VALUE OF t-STATISTICS

TEST	TOTAL TEST	SEQUENCE SUBSCORES	DIFFICULTY	N
Random Responses	2.09	0.86 -0.08	1.63	47
Elementary Algebra	-0.06	0.25 0.38	0.17	760
Logical Relations	-0.17	0.19 0.31	0.00	819

TABLE 4

PERCENT OF RANDOM RESPONSE PATTERNS WITH CRITICAL FIT STATISTICS

$t \geq 1.5$

	N	% of Total (N=47)
Total t	30	64
Sequence	15	32
Subscores	4	9
Difficulty	29	62
One or More Critical Values	37	79
$t < 1.0$	5	11

TABLE 5

SUMMARY OF FIT STATISTICS

	Elementary Algebra	Logical Relationships
Total Sample	822	822
Perfect Scores	3	4
Zero Scores	59	2
Persons in Analysis	760	819
t* > 1.5	104 (14%)	89 (11%)
B1* > 1.5	94 (12%)	67 (8%)
B2* > 1.5	125 (16%)	82 (10%)
B3* > 1.5	79 (10%)	53 (6%)
t, B1, B2, B3 > 1.5	259 (34%)	204 (25%)
B3 > 1.5 and t < 1.5	21 (2%)	24 (3%)
B3 > 1.5 and t < 0	5 (1%)	4 (1%)

* t = Total Test Fit Statistic

B1 = Sequence Between Group Fit Statistic

B2 = Subscore Between Group Fit Statistic

B3 = Difficulty Between Group Fit Statistic

TABLE 6

SUMMARY OF TYPES OF MISFIT*

Elementary Algebra

	t>1.5	B1>1.5	B2>1.5	B3>1.5
Guessing	33	8	29	58
Start-up	1	5	-	7
Content Interaction	3	1	48	5
Speedness	1	-	-	2
Gave up	2	20	7	-
Missed 1 or 2 Very Easy Items	9	1	-	-
Worked 10 or more than Guessed	-	12	-	-

TABLE 7

SCORE DISTRIBUTION OF MISFITTING STUDENTS

Elementary Algebra

Score Range	t>1.5	B1>1.5	B2>1.5	B3>1.5
1 - 9	79	51	62	57
10 - 19	15	31	53	19
20 +	10	12	11	3

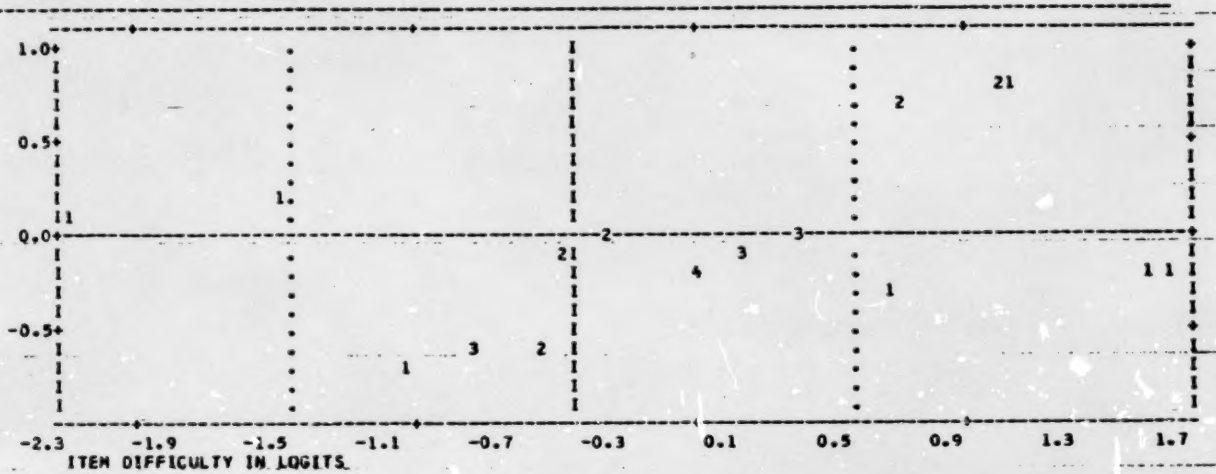
* Individuals with two critical misfit statistics were included under the catagory of greatest misfit.

NJCbSPT ELEMENTARY ALGEBRA TOTAL T GREATER THAN 1.5

PERSON IDENTIFICATION 2 PAGE 71

SUMMARY STATISTICS	
MEASURE	-0.46
SEM	0.40
TOTAL T	2.4
SKIPPING OMITS	
MEASURE	-0.46
SEM	0.40
BETWEEN SUBTEST	
SUBSCORES	23.6
SEQUENCE	8.2
DIFFICULTY	7.0

MEAN RESIDUAL BY DIFFICULTY



NAME	COUNT	SCORE	ABILITY	S.E.	WTHIN	BETWN
SUBSCORES			-0.46	0.40	2.4	23.6
			-0.46	0.40		
SUB1	18.	2.	-2.38	0.77	0.1	
SUB2	12.	10.	1.90	0.80	3.1	

NAME	COUNT	SCORE	ABILITY	S.E.	WTHIN	BETWN
SEQUENCE			-0.46	0.40	2.4	8.2
			-0.46	0.40		
1TEN	10.	4.	-0.83	0.71	-0.3	
2TEN	10.	2.	-1.72	0.81	-1.5	
3TEN	10.	6.	1.00	0.67	2.6	

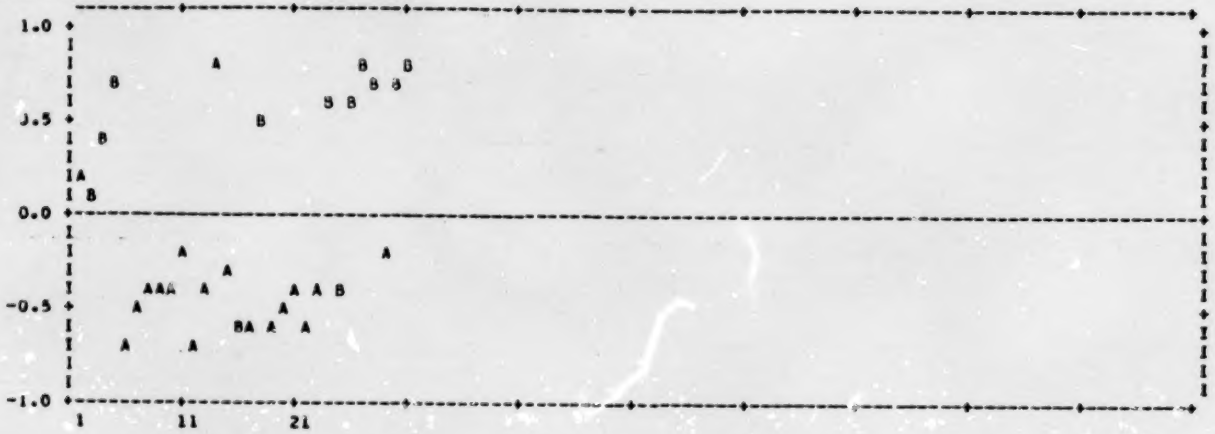
NAME	COUNT	SCORE	ABILITY	S.E.	WTHIN	BETWN
DIFFICULTY			-0.46	0.40	2.4	7.0
			-0.46	0.40		
-.75	6.	2.	-1.96	0.90	0.8	
0.00	8.	2.	-1.46	0.82	-0.7	
0.83	11.	5.	0.14	0.61	1.3	
2.00	5.	3.	1.72	0.92	2.0	

THE HIGH TOTAL T TOGETHER WITH PLOT OF MEAN RESIDUALS BY DIFFICULTY AND THE PATTERN OF ABILITY ESTIMATES FROM THE DIFFICULTY PLOT MIGHT HAVE INDICATED GUESSING IF WE HAD NOT ALSO ANALYZED THE EFFECT OF THE TWO MAJOR CONTENT AREAS. CLEARLY THE STRONG ABILITY IN ONE CONTENT AREA RESULTED IN THE STUDENT ANSWERING MANY DIFFICULT QUESTIONS CORRECTLY MANY MORE THAN WE WOULD HAVE EXPECTED ON THE BASIS OF THE TOTAL TEST SCORE.

NJCBSPT ELEMENTARY ALGEBRA TOTAL T GREATER THAN 1.5

PERSON IDENTIFICATION 2

LOG LOGIT RESIDUALS PLOTTED IN SEQUENCE ORDER



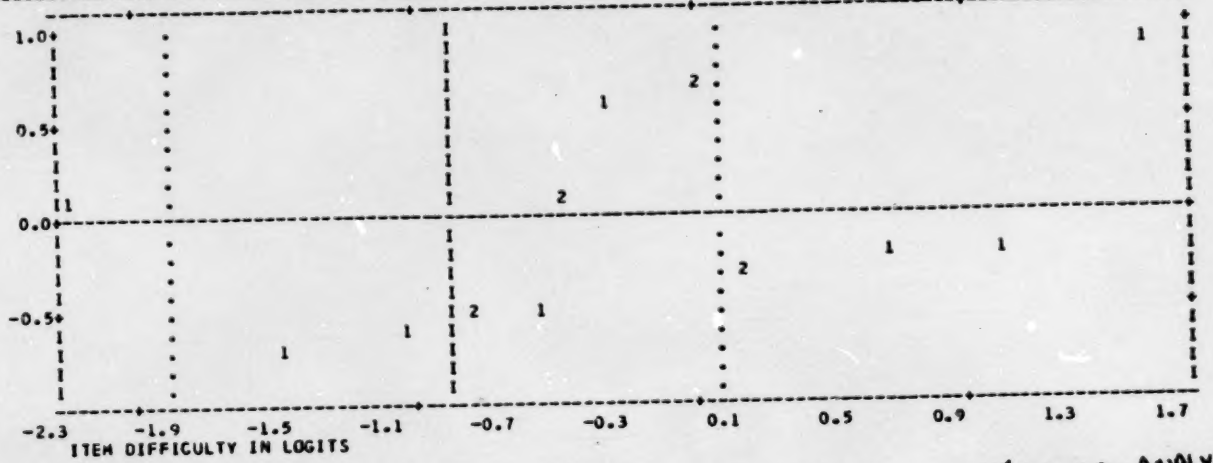
NJCBSPT ELEMENTARY ALGEBRA TOTAL T GREATER THAN 1.5

PERSON IDENTIFICATION 2

SUMMARY STATISTICS 1

MEASURE	-1.59
SEM	0.48
TOTAL T	1.6
SKIPPING UNITS	
MEASURE	-0.91
SEM	0.56
BETWEEN SUBTEST	
SUBSCORES	-1.8
SEQUENCE	-7.8
DIFFICULTY	5.4

MEAN RESIDUAL BY DIFFICULTY



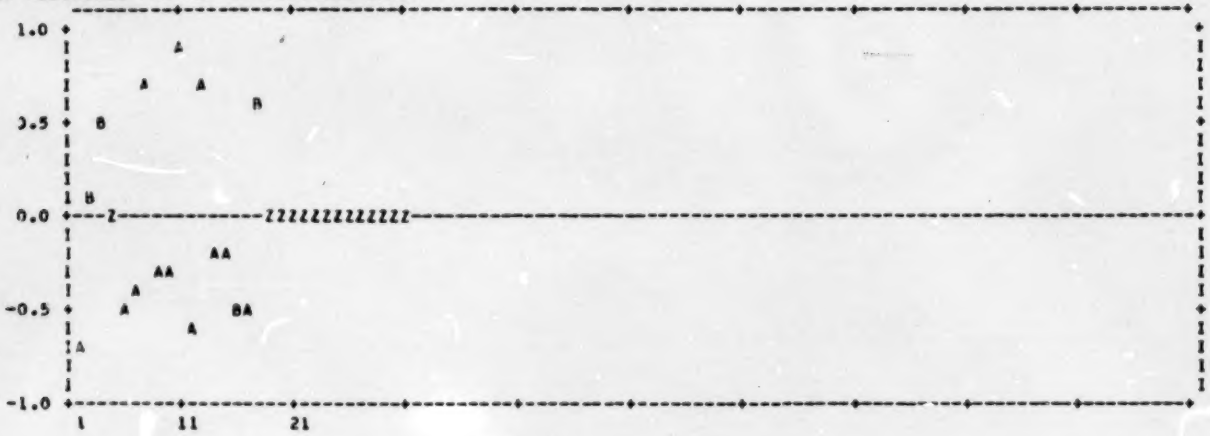
NAME	COUNT	SCORE	ABILITY	S.E.	MTHIN	BETWN
SUBSCORES			-1.59	0.48	1.6	-1.8
			-0.91	0.56		
SUB1	12.	3.	-1.38	0.71	1.6	
SUB2	4.	3.	0.33	1.19	0.5	

NAME	COUNT	SCORE	ABILITY	S.E.	MTHIN	BETWN
SEQUENCE			-1.59	0.48	1.6	-7.8
			-0.91	0.56		
1TEM	9.	4.	-0.70	0.74	1.5	
2TEM	7.	2.	-1.18	0.87	0.6	

NAME	COUNT	SCORE	ABILITY	S.E.	MTHIN	BETWN
DIFFICULTY			-1.59	0.48	1.6	5.4
			-0.91	0.56		
-.75	5.	1.	-2.80	1.15	0.4	
0.00	6.	4.	0.56	0.87	1.9	
0.80	3.	3.	**********		-1.1	
2.00	2.	1.	1.33	1.43	1.7	

IN THE INITIAL ANALYSIS THIS PERSON WAS IDENTIFIED AS A PLODDER BECAUSE OF A HIGH SEQUENCE T. FURTHER INVESTIGATION SHOWED, THAT THE STUDENT HAD NOT ATTEMPTED THE LAST 13 ITEMS ON THE PRETEST. REANALYSIS OMITTING THE UNATTEMPT ITEMS INDICATES THAT THIS IS PROBABLY A CASE OF GUESSING! PARTICULARLY SINCE THE ONE DIFFICULT ITEM ANSWERED CORRECTLY WAS IN SUB1 UNDER SUBSCORES. AND THE INDIVIDUAL MISSED 9 MUCH LESS DIFFICULT ITEMS.

LOG LOGIT RESIDUALS PLOTTED IN SEQUENCE ORDER

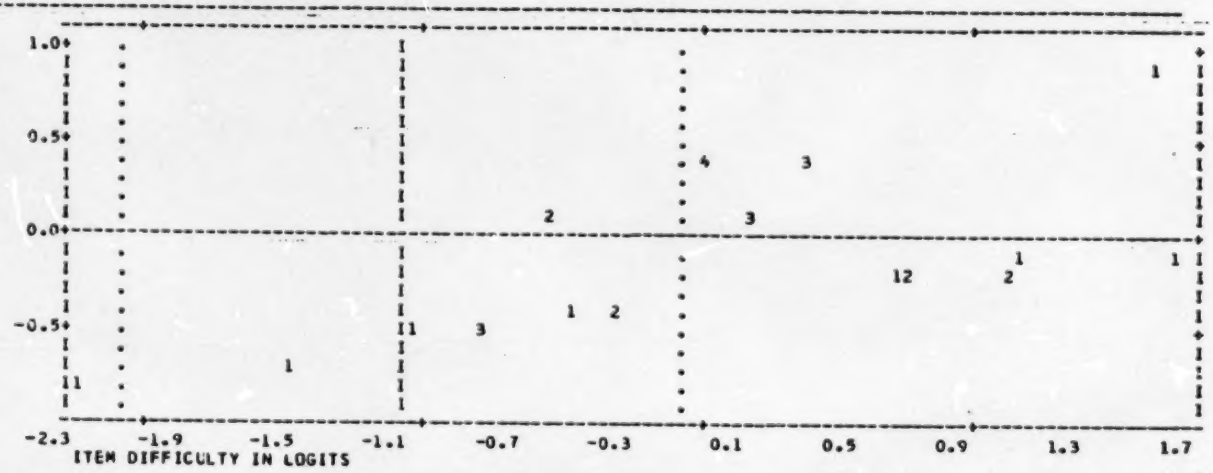


NJCBSPT ELEMENTARY ALGEBRA TOTAL T GREATER THAN 1.5

PERSON IDENTIFICATION 3 PAGE 55

SUMMARY STATISTICS	
MEASURE	-1.16
SEM	0.44
TOTAL T	2.0
SKIPPING UNITS	
MEASURE	-1.10
SEM	0.45
BETWEEN SUBTEST	
SUBSCORES	-2.7
SEQUENCE	2.1
DIFFICULTY	7.7

MEAN RESIDUAL BY DIFFICULTY



NAME	COUNT	SCORE	ABILITY	S.E.	WTHIN	BETWN
SUBSCORES			-1.16	0.44	2.0	-2.7
			-1.10	0.45		
SUB1	17.	6.	-0.75	0.54	2.3	
SUB2	12.	2.	-1.78	0.85	0.6	

NAME	COUNT	SCORE	ABILITY	S.E.	WTHIN	BETWN
SEQUENCE			-1.16	0.44	2.0	2.1
			-1.10	0.45		
1TEN	9.	2.	-1.87	0.88	1.9	
2TEN	10.	2.	-1.72	0.81	0.2	
3TEN	10.	4.	0.12	0.67	1.3	

NAME	COUNT	SCORE	ABILITY	S.E.	WTHIN	BETWN
DIFFICULTY			-1.16	0.44	2.0	7.7
			-1.10	0.45		
-.75	6.				1.3	
0.00	7.	2.	-1.26	0.84	0.0	
0.80	11.	5.	0.14	0.61	1.7	
2.00	5.	1.	-0.10	1.08	1.2	

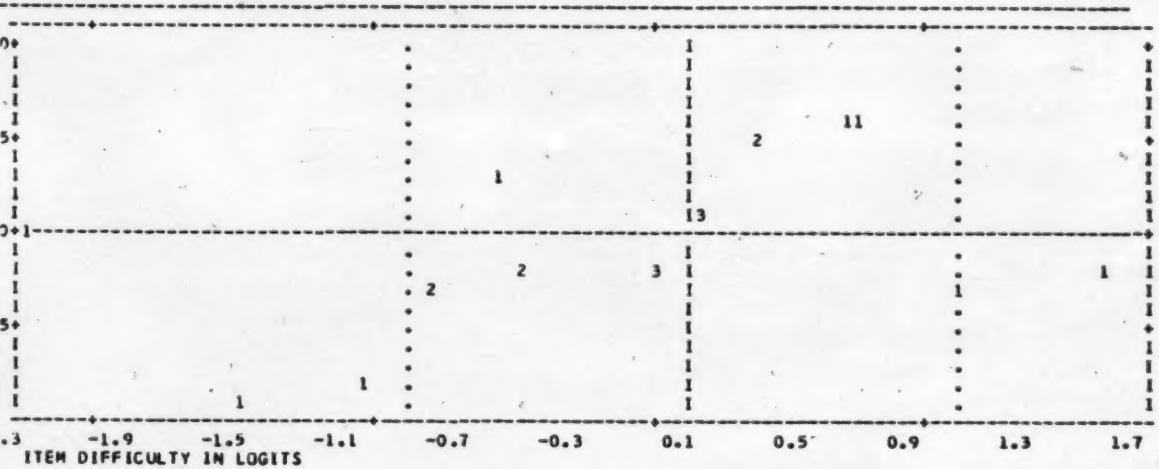
THIS IS A FAIRLY TYPICAL GUESSING PATTERN. THE HIGH TOTAL T AND NO HIGH BETWEEN GROUP T FOR THE DIFFICULTY SUBSETS INDICATE THAT THIS SCORE ALTHOUGH COMMON WAS NO RESULT OF MOSTLY CHANCE AND WAS NOT EARNED.

NJCBSPT ELEMENTARY ALGEBRA TOTAL T GREATER THAN 1.5

PERSON IDENTIFICATION 3 PAGE 7

SUMMARY STATISTICS I MEAN RESIDUAL BY DIFFICULTY

MEASURE	-0.63	I
SEM	0.41	I
TOTAL T	1.9	I
SKIPPING OMITS		I
MEASURE	0.10	I
SEM	0.48	I
BETWEEN SUBTEST		I
SUBSCORES	15.5	I
SEQUENCE	3.8	I
DIFFICULTY	11.1	I



NAME	COUNT	SCORE	ABILITY	S.E.	MNSQ	BETWN
SUBSCORES			-0.63	0.41	1.9	15.5
			0.10	0.48		
SUB1	12.	3.	-1.38	0.71	2.1	
SUB2	8.	8.			0.2	

NAME	COUNT	SCORE	ABILITY	S.E.	MNSQ	BETWN
SEQUENCE			-0.63	0.41	1.9	3.8
			0.10	0.48		
1TEN	10.	5.	-0.33	0.70	0.9	
2TEN	4.	2.	-0.93	0.92	1.6	
3TEN	4.	4.			1.6	

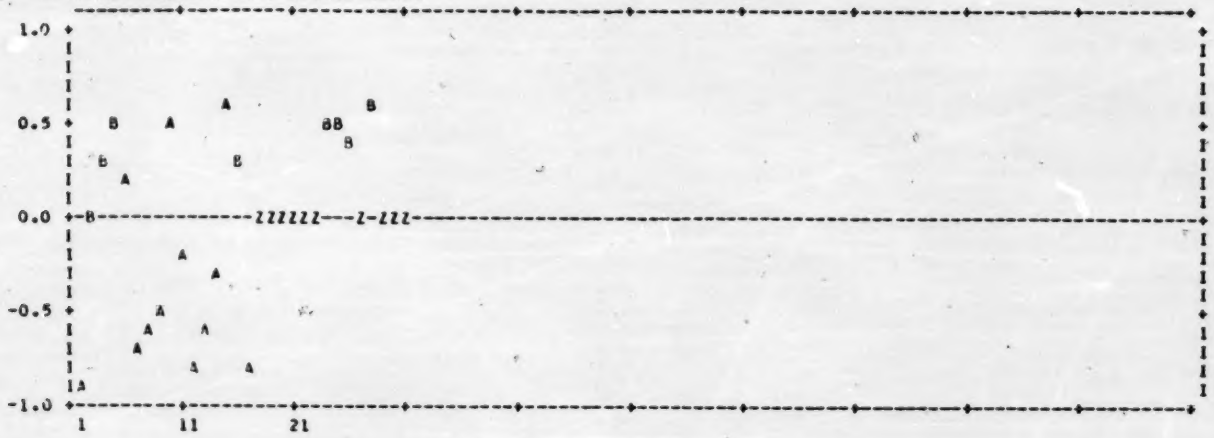
NAME	COUNT	SCORE	ABILITY	S.E.	MNSQ	BETWN
DIFFICULTY			-0.63	0.41	1.9	11.1
			0.10	0.48		
-.75	5.	2.	-1.75	0.95	1.8	
0.50	5.	2.	-0.73	0.92	0.2	
0.90	5.	1.	2.29	1.07	2.3	
2.00	2.	0.			-0.8	

This individual indicates the strong effect that Content interaction has on the total t and the difficulty t. In this analysis the Sequence t is also large but this is due to the fact that omitted items are excluded from analysis and when rushed this student picked only those items which were in his strong content area.

UCBSPT ELEMENTARY ALGEBRA TOTAL T GREATER THAN 1.5

PERSON IDENTIFICATION 3 PAGE 8

LOG LOGIT RESIDUALS PLOTTED IN SEQUENCE ORDER



REFERENCES

Mead, R.J., Assessing The Fit of Data To The Rasch Model. Paper presented at The Annual Conference of the American Educational Research Association, San Francisco, 1976.

Mead, R.J., PANAL Programs, University of Chicago, 1978.

Mead, R.J., and Wright, B.D., The Use of the Rasch Simple Logistic Model To Control The Effects of Random Guessing. Paper presented at the Annual Conference of the American Educational Research Association, New York, 1977.

Smith, R.M. and Mitchell, V.P., Improving The Predictive Ability of Placement Tests Using The Rasch Model For Scoring. Paper presented at the Annual Conference of The National Council on Measurement in Education, San Francisco, 1979.

Wright, B.D., Solving Measurement Problems With The Rasch Model. Journal of Educational Measurement. Vol. 14, No. 2, pp. 97-116.

Wright, B.D., The Rasch Model for Test Construction and Person Measurement. A paper presented at the Fifth Annual Conference and Exhibition on Measurement and Evaluation, Los Angeles, 1978.

Wright, B.D., and Stone, M.H., Best Test Design. Mesa Press, Chicago, Ill., 1979. Ch. 1, pg. 2; Ch 4, pg. 1.

Wright, B.D., and Mead, R.J., BICAL: Calibrating Items and Scales With The Rasch Model. Research Memorandum Number 23A, University of Chicago, 1978.